Brief introduction to mathematical optimization

Department of Mechanics
Faculty of Civil Engineering
Czech Technical University in Prague

Marek Tyburec

Table of contents

[1] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering. Springer, 2 edition, 2006, doi: 10.1007/978-0-387-40065-5

Preliminaries

- Fields: real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, natural numbers $\mathbb{N}$, binary numbers $\mathbb{B}$, integer numbers $\mathbb{Z}$, real symmetric square matrices $\mathbb{S}$
- Superscript denotes the size, subscript additional constraints using element-wise ordering $<$, $\leq$, $\geq$, $>$ or matrix eigenvalue ordering $\prec$, $\preceq$, $\succeq$, $\succ$
- Scalar $a \in \mathbb{Z}_{\leq 0}$
- Vector $\mathbf{x} \in \mathbb{N}^n$ with the i-th component x_i. However, $\mathbf{y}_i \in \mathbb{B}^n$ is a vector indexed by i
- Matrix $\mathbf{Y} \in \mathbb{S}^n_{\succeq 0}$ with the i-th row and j-th column component $Y_{i,j}$
- Positive semidefinite matrix $\mathbf{Y} \in \mathbb{S}^n_{\succeq 0}$
  - $\Leftrightarrow$ $\forall \mathbf{x} \in \mathbb{R}^n : \mathbf{x}^{\mathrm{T}}\mathbf{Y}\mathbf{x} \geq 0$
  - $\Leftrightarrow$ all eigenvalues are real and non-negative
  - $\Leftrightarrow$ $\exists \mathbf{V} \in \mathbb{R}^{n \times n} : \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{Y}$

- Eigenvalue $\lambda \in \mathbb{R}_{\geq 0}$ and eigenvector $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ of a matrix $\mathbf{Y} \in \mathbb{S}^n_{\succeq 0}$ solve the eigenvalue equation

$$\mathbf{Y}\mathbf{x} = \lambda \mathbf{x}$$

- Eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the roots of the characteristic polynomial

$$p(\lambda) = \mathrm{Det}\left(\mathbf{Y} - \lambda \mathbf{I}\right) = \prod_{i=1}^{n} \left(\lambda - \lambda_i\right) = 0$$

- For $\lambda_\bullet$, eigenvector $\mathbf{x}_\bullet$ follows from solving $\left(\mathbf{Y} - \lambda_\bullet \mathbf{I}\right) \mathbf{x}_\bullet = \mathbf{0}$
- Column space/range space/image of a matrix $\mathbf{Y}$ is the span of the column vectors

$$\mathrm{Im}(\mathbf{Y}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Y}\mathbf{x}\}$$

- Nullspace/kernel of a matrix $\mathbf{Y}$ is

$$\mathrm{Ker}(\mathbf{Y}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Y}\mathbf{x} = \mathbf{0}\}$$

- Rank$(\mathbf{Y}) = \mathrm{Dim}\left(\mathrm{Im}(\mathbf{Y})\right)$, Nullity$(\mathbf{Y}) = \mathrm{Dim}\left(\mathrm{Ker}(\mathbf{Y})\right)$

$$\mathrm{Rank}(\mathbf{Y}) + \mathrm{Nullity}(\mathbf{Y}) = n$$

- Function value

$$f(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$$

- Gradient (steepest ascent direction, tangent)

$$\boldsymbol{\nabla} f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

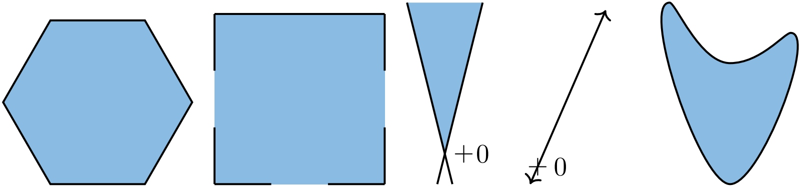
- Hessian (local curvature)

$$\boldsymbol{\nabla}^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Convex sets

Definition (convex set)

A set C is convex if $\forall \mathbf{x}, \mathbf{y} \in C, \theta \in [0,1] : \theta\mathbf{x} + (1-\theta)\mathbf{y} \in C$



- Which of the above sets are convex? Why?

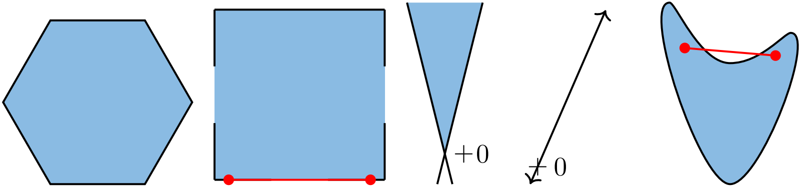
Definition (cone)

A set C is a cone if $\forall \mathbf{x} \in C, \theta \geq 0 : \theta\mathbf{x} \in C$

- What is the relation of convex and conic sets? Are any of the above sets cones?

Definition (convex set)

A set C is convex if $\forall \mathbf{x}, \mathbf{y} \in C, \theta \in [0, 1] : \theta\mathbf{x} + (1 - \theta)\mathbf{y} \in C$



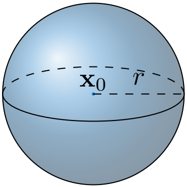
- Which of the above sets are convex? Why?

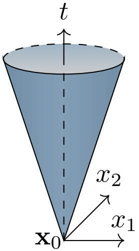
Definition (cone)

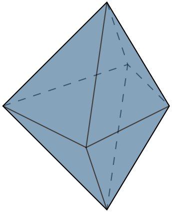
A set C is a cone if $\forall \mathbf{x} \in C, \theta \geq 0 : \theta\mathbf{x} \in C$

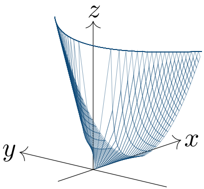
- What is the relation of convex and conic sets? Are any of the above sets cones?

- Examples of convex sets:
  - Empty set $\{\emptyset\}$, singleton $\{\mathbf{x}\}$, the whole space $\mathbb{R}^n$
  - Hyperplanes $\{\mathbf{x} \mid \mathbf{a}^\mathrm{T}\mathbf{x} = b\}$
  - Halfspaces $\{\mathbf{x} \mid \mathbf{a}^\mathrm{T}\mathbf{x} \leq b\}$
  - Norm balls $\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$
  - Norm cones $\{(\mathbf{x}, t) \mid \|\mathbf{x} - \mathbf{x}_0\| \leq t\}$
  - Polyhedra
    $\{\mathbf{x} \mid \mathbf{a}_j^\mathrm{T}\mathbf{x} \leq b_j, j \in \{1, \ldots, m\}, \mathbf{c}_j^\mathrm{T}\mathbf{x} = d_j, j = \{1, \ldots, p\}\}$
  - Positive semidefinite cone $\{\mathbf{X} \in \mathbb{S}^n \mid \mathbf{X} \succeq 0\}$
- How to prove convexity of a set?
- Intersection of convex sets is a convex set









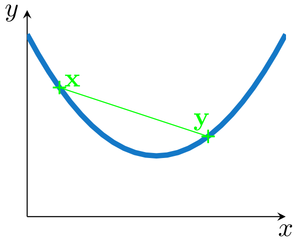
Convex functions

Definition (Jensen's inequality, zeroth-order)

Function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if $\mathrm{Dom}(f)$ is a convex set and $\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f)$ and $\theta \in [0, 1]$ it holds that

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}).$$

- ⇔ first-order:
  $\mathrm{Dom}(f)$ convex, $\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f) : f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{x})$
  - First-order Taylor approximation is a global underestimator
- ⇔ second-order: $\mathrm{Dom}(f)$ convex, $\forall \mathbf{x} \in \mathrm{Dom}(f) : \nabla^2 f(\mathbf{x}) \succeq 0$

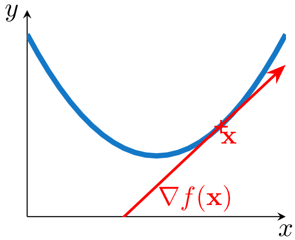


Definition (Jensen's inequality, zeroth-order)

Function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if $\mathrm{Dom}(f)$ is a convex set and $\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f)$ and $\theta \in [0, 1]$ it holds that

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}).$$

- $\Leftrightarrow$ first-order: $\mathrm{Dom}(f)$ convex, $\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f) : f(\mathbf{y}) \geq f(\mathbf{x}) + \boldsymbol{\nabla} f(\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{x})$
  - First-order Taylor approximation is a global underestimator
- $\Leftrightarrow$ second-order: $\mathrm{Dom}(f)$ convex, $\forall \mathbf{x} \in \mathrm{Dom}(f) : \boldsymbol{\nabla}^2 f(\mathbf{x}) \succeq 0$



- Examples of convex functions:
  - Affine and linear functions ($\mathbf{a}^{\mathrm{T}}\mathbf{x} + c$, $\mathbf{a}^{\mathrm{T}}\mathbf{x}$)
  - Quadratic functions $\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x} + \mathbf{q}^{\mathrm{T}}\mathbf{x} + r$ with $\mathbf{P} \in \mathbb{S}^n$, $q \in \mathbb{R}^n$ and $r \in \mathbb{R}$ are convex iff $\mathbf{P} \in \mathbb{S}_{\succeq 0}$, based on second-order condition
  - Exponential functions e^{ax}, on $\mathbb{R}$ with $a \in \mathbb{R}$
  - Power functions x^a, on $\mathbb{R}_{>0}$ with $a \in (-\infty, 0] \cup [1, \infty)$
  - Norms
  - Max function $\max\{x_1, \dots, x_n\}$

Lemma

Assume an unconstrained optimization problem $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$ *with* $f(\mathbf{x})$ *convex and differentiable. Then, any point* $\overline{\mathbf{x}}$ *satisfying* $\boldsymbol{\nabla} f(\overline{\mathbf{x}}) = \mathbf{0}$ *is a global minimizer.*

Proof.

Using the first-order definition of convexity, we have

$$\forall \mathbf{x}, \mathbf{y} : f(\mathbf{y}) \geq f(\mathbf{x}) + \boldsymbol{\nabla} f(\overline{\mathbf{x}})^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{x}})$$

Since $\boldsymbol{\nabla} f(\overline{\mathbf{x}}) = 0$, we receive

$$\forall \mathbf{x}, \mathbf{y} : f(\mathbf{y}) \geq f(\mathbf{x}),$$

concluding that $\overline{\mathbf{x}}$ is indeed a global minimizer. □

- Strictly convex function $\longrightarrow$ unique minimizer

Definition (quasiconvex function, zeroth-order)

Function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is quasiconvex if its domain and all sublevel sets $S_\alpha = \{\mathbf{x} \in \mathrm{Dom}(f) \mid f(\mathbf{x}) \leq \alpha\}$ for $\alpha \in \mathbb{R}$ are convex

$\Leftrightarrow$ first-order: $\mathrm{Dom}(f)$ convex and
$\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f) : f(\mathbf{y}) \leq f(\mathbf{x}) \Rightarrow \boldsymbol{\nabla} f(\mathbf{x})^\mathrm{T}(\mathbf{y} - \mathbf{x}) \leq 0$

- When $\boldsymbol{\nabla} f(\mathbf{x}) \neq \mathbf{0}$, $\boldsymbol{\nabla} f(\mathbf{x})$ defines a supporting hyperplane to the sublevel set $\{\mathbf{y} \mid f(\mathbf{y}) \leq \mathbf{x}\}$
- $\boldsymbol{\nabla} f(\mathbf{x}) = \mathbf{0}$ does not imply global optimality

$\Leftrightarrow$ second-order: $\mathrm{Dom}(f)$ convex and
$\forall \mathbf{x}, \mathbf{y} \in \mathrm{Dom}(f) : \mathbf{y}^\mathrm{T} \boldsymbol{\nabla} f(\mathbf{x}) = 0 \Rightarrow \mathbf{y}^\mathrm{T} \boldsymbol{\nabla}^2 f(\mathbf{x}) \mathbf{y} \geq 0$

- In 1D, at any point with zero slope, the second-derivative is non-negative

Convex optimization problems

- Formalization of optimization problems

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \; & f(\mathbf{x}) && \text{objective/cost function} \\ \text{s.t. } & g_i(\mathbf{x}) \le 0, \forall i \in \{1,\dots,m\}, && \text{inequality constraints} \\ & h_j(\mathbf{x}) = 0, \forall j \in \{1,\dots,\ell\}, && \text{equality constraints} \end{aligned}$$

- Optimization variable $\mathbf{x}$
- When $i \in \{\emptyset\}$ and $j \in \{\emptyset\}$, the problem is called unconstrained
- Domain of the opt. problem $\mathcal{D} = \bigcap_{i=1}^{m} \mathrm{Dom}(g_i) \cap \bigcap_{j=1}^{\ell} \mathrm{Dom}(h_j)$
- $\mathbf{x} \in \mathcal{D}$ is called a feasible point if it satisfies all the constraints, infeasible otherwise
- Feasible set: the set of all feasible points
- Optimal value $f^* = \inf\{f(\mathbf{x}) \mid g_i(\mathbf{x}) \le 0, i \in \{1,\dots,m\}, h_j(\mathbf{x}) = 0, j \in \{1,\dots,\ell\}\}$, $f^* = \infty$ when infeasible and $f^* = -\infty$ when unbounded from below

- $\mathbf{x}^*$ is optimal point if it is feasible and $f(\mathbf{x}^*) = f^*$
- Optimal set $\mathcal{X}_{\text{opt}} = \{\mathbf{x} \mid g_i(\mathbf{x}) \leq 0, i \in \{1, \ldots, m\}, h_j(\mathbf{x}) = 0, j \in \{1, \ldots, \ell\}, f(\mathbf{x}) = f^*\}$
- When $\mathcal{X} \neq \{\emptyset\}$, the problem is solvable and the optimum value is attained
- Feasible point $\mathbf{x}$ with $f(\mathbf{x}) \leq f^* + \varepsilon$ is ε-suboptimal
- Feasible point $\mathbf{x}$ is locally-optimal if $\exists r > 0$ such that $f(\mathbf{x}) = \inf\{f(\mathbf{z}) \mid g_i(\mathbf{z}) \leq 0, i \in \{1, \ldots, m\}, h_j(\mathbf{z}) = 0, j \in \{1, \ldots, \ell\}, \|\mathbf{x} - \mathbf{z}\| \leq r\}$
- Inequality constraint is active when $g_i(\mathbf{x}) = 0$ and inactive otherwise
- Redundant constraint does not change the feasible set

- The optimization problem

$$\begin{aligned}\min_{\mathbf{x}\in\mathbb{R}^n}\ & f(\mathbf{x})\\ \text{s.t. } & g_i(\mathbf{x}) \leq 0, \forall i \in \{1,\dots,m\},\\ & h_j(\mathbf{x}) = 0, \forall j \in \{1,\dots,\ell\},\end{aligned}$$

  is convex if $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are convex and and $h_j(\mathbf{x})$ is affine
- The feasible set is convex as it is an intersection of convex domains
- For $f(\mathbf{x})$ quasiconvex instead of convex, the problem is called quasiconvex
- All ε-sublevel sets are convex $\Rightarrow$ ε-suboptimal sets are convex $\Rightarrow$ optimal set is convex
- Opt. problem is equivalent to

$$\begin{aligned}\max_{\mathbf{x}\in\mathbb{R}^n}\ & \hat{f}(\mathbf{x})\\ \text{s.t. } & g_i(\mathbf{x}) \leq 0, \forall i \in \{1,\dots,m\},\\ & h_j(\mathbf{x}) = 0, \forall j \in \{1,\dots,\ell\},\end{aligned}$$

  with $\hat{f}(\mathbf{x}) = -f(\mathbf{x})$ concave

Lemma

Let $\mathcal{P}$ be a convex optimization problem. Then, any locally optimal point $\mathbf{x}^*$ is also globally optimal.

Proof.

Let $\hat{\mathbf{x}}$ be feasible and locally optimal, i.e.,

$$f(\hat{\mathbf{x}}) = \inf\left\{f(\mathbf{x}) \mid \mathbf{x} \text{ feasible}, \|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq R\right\}, \text{where } R > 0.$$

For $\hat{\mathbf{x}}$ not globally optimal, there is an $\mathbf{y}$ such that $f(\mathbf{y}) < f(\hat{\mathbf{x}})$ and $\|\mathbf{y}-\hat{\mathbf{x}}\|_2 > R$.
Further, we set $\mathbf{z} = (1-\theta)\hat{\mathbf{x}} + \theta\mathbf{y}$ with $\theta = \frac{R}{2\|\mathbf{y}-\hat{\mathbf{x}}\|_2}$. Then, $\|\mathbf{z}-\hat{\mathbf{x}}\|_2 = \frac{R}{2} < R$ and $\mathbf{z}$ is feasible by convexity of the feasible set.
By convexity of $f(\mathbf{x})$, we also have

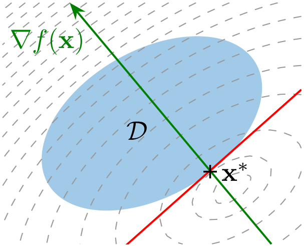
$$f(\mathbf{z}) \leq (1-\theta)f(\hat{\mathbf{x}}) + \theta f(\mathbf{y}) < f(\hat{\mathbf{x}})$$

which contradicts the local optimality assumption. □

Lemma

Let $f(\mathbf{x})$ *be differentiable and convex. Then,* $\mathbf{x}^*$ *is optimal iff* $\mathbf{x}^* \in \mathcal{D}$ *and*

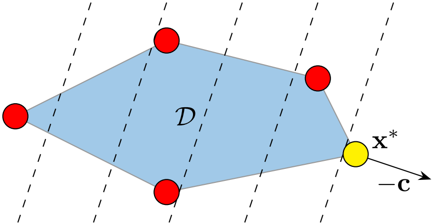
$$\forall \mathbf{y} \in \mathcal{D} : \nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{y} - \mathbf{x}^*) \geq 0$$



- Standard form

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \ & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{s.t. } & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

- Linear objective function and linear inequality constraints
- **Q**: How to write equality constraint in this form?
- **Q**: What is the shape of the feasible set?



- Standard form

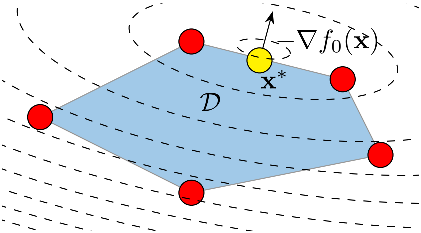
$$\min_{\mathbf{x}\in\mathbb{R}^n} \; \mathbf{c}^{\mathrm{T}}\mathbf{x}$$
$$\text{s.t.} \; \mathbf{A}\mathbf{x} \leq \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0}$$

- Linear objective function and linear inequality constraints
- **Q**: How to write equality constraint in this form?
- **Q**: What is the shape of the feasible set? Convex polyhedron
- If bounded and feasible, optimum value attained at the boundary of $\mathcal{D}$
- Vertices: basic feasible solutions, intersection of d inequality constraints
- Simplex algorithm

- Formulation

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x} + \mathbf{q}^{\mathrm{T}}\mathbf{x}$$
$$\text{s.t. } \mathbf{G}\mathbf{x} \leq \mathbf{h},$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{P} \in \mathbb{S}^n_{\succeq 0}$,$\mathbf{G} \in \mathbb{R}^{m\times n}$ and $\mathbf{a} \in \mathbb{R}^{p\times n}$



- Formulation

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \ & \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x} + \mathbf{q}^{\mathrm{T}}\mathbf{x} \\ \text{s.t. } & \mathbf{G}\mathbf{x} \leq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

  where $\mathbf{P} \in \mathbb{S}^n_{\succeq 0}$,$\mathbf{G} \in \mathbb{R}^{m\times n}$ and $\mathbf{a} \in \mathbb{R}^{p\times n}$
- We have $\nabla f(\mathbf{x}) = \mathbf{P}\mathbf{x} + \mathbf{q}$ and $\nabla^2 f(\mathbf{x}) = \mathbf{P}$
  - $\mathbf{P} \in \mathbb{S}^n_{\succeq 0}$: convex problem
  - $\mathbf{P} = \mathbf{0}$: linear programming problem
  - $\mathbf{P}$ indefinite: $\mathcal{NP}$-hard
- Addition of convex quadratic constraints

$$\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{P}_i\mathbf{x} + \mathbf{q}_i^{\mathrm{T}}\mathbf{x} + r_i \leq 0$$

  $\Longrightarrow$ quadratically constrained quadratic program

- Formulation

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \ & \mathbf{f}^{\mathrm{T}}\mathbf{x} \\ \text{s.t.} \ & \|\mathbf{A}_i\mathbf{x}+\mathbf{b}_i\|_2 \leq \mathbf{c}_i^{\mathrm{T}}\mathbf{x}+d_i \\ & \mathbf{F}\mathbf{x}=\mathbf{g} \end{aligned}$$

- With $\mathbf{A}_i \in \mathbb{R}^{k\times n}$, $\left(\mathbf{A}_i\mathbf{x}+\mathbf{b}_i, \mathbf{c}_i^{\mathrm{T}}\mathbf{x}+d_i\right)$ lies in the second-order cone in $\mathbb{R}^{k+1}$
- With $\mathbf{A}_i = \mathbf{0}$, reduction to linear programming
- With $\mathbf{c}_i = \mathbf{0}$, reduction to quadratically constrained quadratic programming

- Formulation

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \ & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{s.t.} \ & \mathbf{A}_0 + \sum_{i=1}^{n} x_i \mathbf{A}_i \preceq 0 \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- Reduction from linear programming with $\mathbf{A}_i$ diagonal
- Reduction from second-order cone programming with

$$\|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^{\mathrm{T}}\mathbf{x} + d_i \iff \begin{pmatrix} (\mathbf{c}_i^{\mathrm{T}}\mathbf{x} + d_i)\mathbf{I} & \mathbf{A}_i\mathbf{x} + \mathbf{b}_i \\ (\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)^{\mathrm{T}} & \mathbf{c}_i^{\mathrm{T}}\mathbf{x} + d_i \end{pmatrix} \succeq 0$$

Duality

- Optimization problem $\mathcal{P}$

$$\begin{aligned} p^* = \min_{\mathbf{x}\in\mathbb{R}^n} \ & f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) \leq 0, \quad i \in \{1, \ldots, m\} \\ & h_j(\mathbf{x}) = 0, \quad j \in \{1, \ldots, \ell\} \end{aligned}$$

- Augment the objective function with a weighted sum of constraint functions

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \nu_j h_j(\mathbf{x})$$

- $\text{Dom}(\mathcal{L}) = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^\ell$
- $\boldsymbol{\lambda}$: Lagrange multipliers associated with inequality constraints
- $\boldsymbol{\nu}$: Lagrange multipliers associated with equality constraints
- $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$ are the dual variables of the problem

- Define a Lagrangian dual function

$$d(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$
$$= \inf_{\mathbf{x} \in \mathcal{D}} \left(f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \nu_j h_j(\mathbf{x}) \right)$$

- When $d(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is unbounded below in $\mathbf{x}$, we extend its value to $-\infty$
- $d(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is always convex, even when the original problem $\mathcal{P}$ is not
- For any $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\nu}$, we have $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$
  - Let $\tilde{\mathbf{x}}$ be feasible to $\mathcal{P}$ and let $\boldsymbol{\lambda} \geq \mathbf{0}$
  - Then, we have $\sum_{i=1}^{m} \lambda_i g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^{\ell} \nu_j h_j(\tilde{\mathbf{x}}) \leq 0$
  - Thus,
    $\mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i g_i(\tilde{\mathbf{x}}) + \sum_{j=1}^{\ell} \nu_j h_j(\tilde{\mathbf{x}}) \leq f(\tilde{\mathbf{x}})$
  - Finally, $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f(\tilde{\mathbf{x}})$
- If $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$, the pair $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is called dual feasible

- Consider linear program in the equality form

$$\begin{aligned}\min_{\mathbf{x}\in\mathbb{R}^n} \ & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{s.t. } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}\end{aligned}$$

- The Lagrangian evaluates as

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \mathbf{c}^{\mathrm{T}}\mathbf{x} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\nu}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= -\boldsymbol{\nu}^{\mathrm{T}}\mathbf{b} + \left(\mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\nu}\right)^{\mathrm{T}}\mathbf{x}\end{aligned}$$

- Dual function

$$\begin{aligned}d(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= \inf_{\mathbf{x}} \left[-\boldsymbol{\nu}^{\mathrm{T}}\mathbf{b} + \left(\mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\nu}\right)^{\mathrm{T}}\mathbf{x}\right]\end{aligned}$$

$$d(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^{\mathrm{T}}\boldsymbol{\nu} & \text{if } \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\nu} = \mathbf{0} \\ -\infty & \text{otherwise}\end{cases}$$

- We have that $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$ for any $\boldsymbol{\lambda} \geq \mathbf{0}$: what is the best lower bound?
- Dual optimization problem

$$d^* = \max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^\ell} d(\boldsymbol{\lambda}, \boldsymbol{\nu}) \quad \text{s.t. } \boldsymbol{\lambda} \geq \mathbf{0}$$

- Convex optimization problem: why?
- Example: linear programming

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^\ell} -\mathbf{b}^{\mathrm{T}} \boldsymbol{\nu} \quad \text{s.t. } \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^{\mathrm{T}} \boldsymbol{\nu} = \mathbf{0}$$

- In general, we have weak duality, i.e., $d^* \leq p^*$

- Strong duality $d^* = p^*$ does not hold in general
- It *usually* holds for convex problems

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \ & f(\mathbf{x}) \\ \text{s.t.} \ & g_i(\mathbf{x}) \leq 0, i \in 1, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- Conditions under which strong duality holds: constraint qualifications
- Slater's condition: there exists $\tilde{\mathbf{x}}$ strictly feasible, that is $\tilde{\mathbf{x}} \in \text{Relint}(\mathcal{D})$, or

$$\tilde{\mathbf{x}} \in \{\mathbf{x} \mid \forall i \in \{1, \ldots, m\} : g_i(\mathbf{x}) < 0, \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

- $f(\mathbf{x}) - d(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is a certificate of ε-suboptimality
- Let $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$, and $\boldsymbol{\nu}^*$ be primal and dual optimal and let $d^* = p^*$. Then,

$$f(\mathbf{x}^*) = d(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \inf_{\mathbf{x}\in\mathbb{R}^n} \left(f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \nu_j^* h_j(\mathbf{x}) \right)$$

$$\leq f(\mathbf{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^{\ell} \nu_j^* h_j^*(\mathbf{x}^*)}_{\leq 0} \leq f(\mathbf{x}^*)$$

- Consequently,
  - $\forall i \in \{1, \dots, m\} : \lambda_i^* g_i(\mathbf{x}^*) = 0$ (complementary slackness)
  - $\mathbf{x}^*$ is a minimizer of $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$

- Assume that g_i and h_i are differentiable and let $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ be optimal primal and dual points with zero optimality gap, we have

$$\begin{aligned}
\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{\ell} \nu_j^* \nabla h_j(\mathbf{x}^*) &= \mathbf{0} \text{ (stationarity of } \mathcal{L}) \\
\forall i \in \{1, \dots, m\} : g_i(\mathbf{x}^*) &\leq 0 \text{ (primal feas.)} \\
\forall j \in \{1, \dots, \ell\} : h_j(\mathbf{x}^*) &= 0 \text{ (primal feas.)} \\
\boldsymbol{\lambda}^* &\geq \mathbf{0} \text{ (dual feas.)} \\
\forall i \in \{1, \dots, m\} : \lambda_i^* g_i(\mathbf{x}^*) &= 0 \text{ (compl. slackness)}
\end{aligned}$$

- For primal convex, KKT conditions are sufficient for the points to be primal and dual optimal
- For convex problems satisfying Slater constraint qualification, KKT are sufficient and necessary conditions of optimality

- S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, doi: 10.1017/cbo9780511804441
- A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Society for Industrial and Applied Mathematics, 2001, doi: 10.1137/1.9780898718829